Osteoarthritis and Cartilage 20 (2012) 638–645

Osteoarthritis and Cartilage

The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it

F.M. Impellizzeri †[a], A.F. Mannion †‡*[a], F.D. Naal §, O. Hersche §, M. Leunig §‖

† Department of Research and Development, Schulthess Clinic, Zurich, Switzerland
‡ Department of Rheumatology and Institute of Physical Medicine, University of Zurich, University Hospital, Switzerland
§ Department of Orthopaedic Surgery, Schulthess Clinic, Zurich, Switzerland
‖ Department of Orthopaedic Surgery, University of Berne, Berne, Switzerland

ARTICLE INFO

Article history:
Received 30 June 2011
Accepted 21 March 2012

Keywords:
Femoroacetabular impingement
Patient-reported outcome
Surgical success
Acceptability of symptoms
Global outcome

SUMMARY

Objective: To evaluate the proportion of "successes" after surgery for femoroacetabular impingement (FAI) using different external criteria, "feeling better" and "feeling good", and to determine the corresponding cut-off scores indicating "success" for the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (0–10-point response scale), Oxford Hip Score (OHS) and EuroQoL-5D (EQ-5D and EQ-VAS).

Design: Prospective, observational study based in an orthopaedic hospital. Ninety-nine consecutive patients with FAI completed the questionnaires before and 6 months after surgery (arthroscopy or mini-open surgical dislocation). Patient-ratings of change in state ("feeling better") were assessed using a global treatment outcome (GTO) item. Acceptability of the current health state was assessed using the symptom-specific well-being (SSWB) item. Cut-off (threshold) scores for the different instruments indicating the minimal clinically important change (MCIC) and acceptable symptom state were calculated using Receiver Operating Characteristics (ROC) analyses.

Results: Significant improvements in all scores ($P < 0.001$) were recorded 6 months after surgery. The proportion of good outcomes measured with GTO was 60%; 55% of patients reported having achieved an acceptable symptom state. The MCIC scores for improvement were ≥6 for the OHS (0–48 total score range), ≥15 for EQ-VAS, ≥0.16 for EQ-5D index, and ≥22 for the WOMAC-total score (0–100 total score range); absolute scores of ≥40, ≥80, ≥0.682 and ≤8, respectively, were associated with an acceptable symptom state.

Conclusions: The results show that feeling better does not always equate to feeling good, and that improvements in outcome scores, even large, do not necessarily indicate acceptability of the current state. The cut-off values may help in the interpretation of trial results and individual change-scores recorded in clinical practice.

Introduction

Femoroacetabular impingement (FAI) occurs when there is a repetitive impact of the femoral head/neck against the acetabular labrum and/or its adjacent cartilage. It is a common source of hip pain and dysfunction in young adults, especially those who are physically active[1,2]. In recent years, convincing evidence has emerged to support the hypothesis that FAI is an etiologic factor in the pathophysiology of osteoarthritis[3].

There is a growing body of literature describing the etiology, diagnosis and treatment of FAI[3–5]. It is generally agreed that surgery to preserve the joint should be taken into consideration at an early stage, since the outcome is usually worse in the presence of significant cartilage damage[4]. Joint-preserving surgery aims to create impingement-free motion through restoration of a more normal morphology in order to alleviate symptoms and to prevent or delay the progression of irreversible degenerative changes[6,7].

The use of patient-oriented measures in the assessment of treatment outcome has now become a widespread phenomenon[8]. A recent systematic review reported mean improvements in pain of between 25% and 100% after surgery for FAI, with 68–100% of

* Address correspondence and reprint requests to: A.F. Mannion, Department of Research and Development, Schulthess Clinic, Lengghalde 2, 8008 Zurich, Switzerland. Tel: 41-44-385-7584; Fax: 41-33-385-7590.
E-mail address: anne.mannion@kws.ch (A.F. Mannion).
[a] The first two authors contributed equally to the work.

1063-4584/$ – see front matter © 2012 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2012.03.019

patients being satisfied with the procedure or reporting improvement in symptoms[4]. Another review reported that in 10 out of 12 Level IV case series, at least 75% patients declared the outcome successful[9]. However, both reviews indicated that the varying and sometimes non-validated outcome measures used in the literature provided a challenge in assessing the true value of treatment, especially on an individual level.

An important and increasingly recognised issue in the field of outcomes research is whether an observed change in an individual's scores for symptoms or function is merely caused by measurement error or constitutes a real change, and whether that change is also clinically relevant[10]. In the literature, the latter is commonly referred to as the "minimal clinically important change" (MCIC)[11]. Differences exist in the methods used to quantify the MCIC, with the most clinically robust of these relying on the comparison of change scores with an external criterion indicating the patient's perception of overall improvement[12,13]. Other concepts addressing clinically relevant treatment success include the symptom-specific well-being (SSWB) and patient acceptable symptom state (PASS)[14], which indicate the score beyond which patients consider themselves well or consider their health state to be acceptable.

An optimal definition of success may be difficult to ascertain in the case of FAI, where the indications for surgery can be a mixture of both symptom relief and joint-preservation, and it is currently unknown how well these individual markers of success perform in this pathology. Few studies have quantified either the responsiveness (i.e., the ability to detect change) or the MCIC, PASS, or SSWB for the instruments used in the assessment of outcome in FAI[15]. Determination of such values would assist in interpreting scores/change scores at the individual level and in establishing realistic expectations regarding the likely outcome of treatment, to counsel patients preoperatively.

In the present study, we examined the association between different expressions of "success" in relation to the early outcome of surgical treatment for FAI. The concepts under investigation included both improvement (feeling better) and acceptability of the current state (feeling good)[16]. The proportion of successes measured using these different external criteria was examined, as were the corresponding cut-off scores associated with a "successful result" for two commonly used hip outcome instruments (WOMAC and Oxford-12) and a generic health-related quality of life questionnaire (EuroQoL-5D).

Methods

Patients, surgery, and study design

One hundred and seventy-two consecutive patients [86 (50%) men and 86 (50%) women] undergoing either arthroscopic surgery with labral preservation (surgeon 1) or limited anterolateral open surgery with labral resection (surgeon 2) for FAI in our hospital between July 2008 and April 2010 were eligible for participation in the study. Inclusion criteria included: cam, localized pincer, or mild-moderate mixed impingement in hips with at most early-OA ($<1^{\circ}$, radiographically assessed); arthroscopic or mini-open osteochondroplasty as the foreseen surgical intervention; and a good understanding of written German. Exclusion criteria included: symptomatic hip dysplasia (lateral centre edge angle $<20^{\circ}$); combined FAI with a global over-coverage (i.e., coxa profunda or protrusio) and extra-articular impingement; and hips requiring cartilage repair techniques (for these, the postoperative restrictions and rehabilitation were more extensive compared with the included surgical techniques, for which full weight-bearing was permitted 2 weeks postoperatively).

A few weeks before admission for surgery, patients were mailed an invitation to participate, an information letter, a set of questionnaires, and an informed consent form. They were requested to complete the questionnaire booklet and send it back the week before admission. After 6 months, a second questionnaire booklet was mailed to those that had returned a preoperative questionnaire, with the request to complete it and return it using the stamped addressed envelope enclosed. One hundred and twenty-eight patients (75%) returned the preoperative questionnaire. One hundred and two (75%) of these also returned a 6-month follow-up questionnaire. Their sociodemographic characteristics are shown in Table I. The study was approved by the local ethical committee and all patients gave their written informed consent to participate.

Questionnaires

The preoperative questionnaire booklet contained the following

(1) the Euro Quality of Life-Five Dimensions index (EQ-5D) and the Euro Quality of Life "visual analogue scale" (EQ-VAS) for health-related quality of life. The EQ-5D is a standardized instrument used as a measure of health outcome[17,18]. It comprises five single items – mobility, self-care, usual activities, pain/discomfort and anxiety/depression – each rated on a 3-point response scale. Unweighted summary index scores (ranging from −0.59 to 1) were computed[19]. The EQ-VAS is a 0–100 scale for assessing current health-related quality of life.
(2) the Oxford Hip Score (OHS). The OHS consists of 12 questions asking patients to describe their hip pain and function during

Table I
Baseline sociodemographic data of the patients (values are mean (SD) unless otherwise stated)

	All patients	Female	Male
Number (%)	102 (100%)	57 (56%)	45 (44%)
Age (yrs)	35.9 (11.5)	37.7 (12.2)	33.7 (10.3)
BMI (kg m^{-2})	23.5 (4.3)	22.4 (3.0)	25.0 (5.2)
Living condition			
Rural	61%	67%	53%
City	39%	33%	47%
Education			
Primary school	12%	10%	13%
High school	5%	9%	0%
Professional school	68%	72%	65%
University	15%	9%	22%
Civil status			
Married	47%	51%	42%
Single	49%	42%	58%
Divorced/separated	4%	7%	0%
Work activity			
Full time	51%	33%	73%
Part time	25%	40%	7%
Student	5%	4%	7%
Homemaker	8%	14%	0%
Retired	1%	0%	2%
Disabled	5%	5%	4%
Other	5%	4%	7%
Operated hip			
Right	52%	44%	62%
Left	48%	56%	38%
Contralateral side			
Involved but not operated	36%	35%	38%
Involved and operated	8%	5%	11%
Not involved	56%	60%	51%
Comorbidities			
None	60%	54%	67%
One	29%	32%	27%
Two	9%	12%	4%
More than three	2%	2%	2%

the past 4 weeks[20,21]. Each item uses a response scale with values from 0 to 4. The responses to the 12 questions are summed to give a total score ranging from 0 to 48, where 0 is the worst possible score and 48 is the best score[22].

(3) the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). This is a 24-item self-administered, disease-specific instrument for assessing pain, stiffness and physical function in osteoarthritis patients[23,24]. In the present study we used the version with the 0–10 numeric scale. Five items address pain, two items address stiffness and 17 items assess physical function. The scores for each subscale and the total score were converted to a 0–100 scale with higher scores indicating a worse health state.

At the 6-month follow-up, the above questionnaires were supplemented with a range of single item questions aimed at assessing the success of surgery, as follows:

Assessment of change/improvement

(1) global treatment outcome (GTO) adapted from Mannion et al[25]. "How much did the operation help your hip problem?" (helped a lot, helped, helped only little, didn't help, made things worse)

Assessment of acceptability of current health state

(2) Symptom-specific well-being (SSWB)[26] "If you had to spend the rest of your life with the symptoms you have now, how would you feel about it?" (very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied)

For some analyses, the 5-point measures of "success" were collapsed to provide dichotomous (good/poor; acceptable/not acceptable) outcome variables. As previously published for patients undergoing elective surgery[25], the first two categories of the 5-point scale for GTO i.e., "operation helped/helped a lot" were taken to represent the cut-off for a "good" outcome. For dichotomising the SSWB responses into "acceptable" vs "not acceptable" states, we performed receiver operating characteristics (ROC) analysis of the data from an initial subgroup of patients who had completed both the 5-point SSWB and the PASS[14]. The PASS asks "Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?" with the answers "yes" or "no". These dichotomous responses served as the external criterion in ROC analysis and the results showed that the top three categories (very satisfied, somewhat satisfied, neither/nor) gave the best indication of an "acceptable state" (90% sensitivity, 90% specificity; ROC area under the curve (AUC) 0.93, 95% CI 0.85 to 0.98; unpublished observations).

Statistical analysis

Analyses were performed using the data of the 102 patients who returned a 6-month follow-up questionnaire. Missing data were dealt with using the corresponding missing rules for the given questionnaire.

Descriptive data indicate the proportions (%) of responses in each global outcome category, and the % categorized as "good" and "poor" after dichotomisation of the 5-point scales. The significance of the difference in an instrument's scale/subscale scores at 6 months compared with baseline was analysed using a 2-way analysis of variance with repeated measures (with gender as the between-factor and time of assessment the within-factor). Assumption of sphericity was examined using the Mauchly's test with Huynh-Feldt correction where needed. Responsiveness was given by the standardized response means [SRM=(posttest mean−pretest mean)/SD changes][27] for the good and poor outcome groups separately, in order to see whether the questionnaires had the ability to differentiate between different global outcomes[28].

The sensitivity and specificity of the WOMAC, OHS and EQ-5D in identifying a "good" outcome, was examined using the ROC method[12]. This is considered analogous to evaluating a diagnostic test, in which the instrument is the diagnostic test and the global outcome represents the gold standard[12]. The ROC curve combines information on sensitivity and specificity for detecting a good global outcome and comprises a plot of 'true-positive rate' (sensitivity) vs 'false positive rate' (1-specificity) for each of several possible cut-off points in change score. The ROC AUC indicates the probability of correctly discriminating between patients with a "good" and a "poor" result, based on the change in instrument scores and can range from 0.5 (no accuracy in discriminating) to 1.0 (perfect accuracy in discriminating). The ROC curve was also used to indicate the cut-off change score for indicating a "good' outcome (i.e., the MCIC for improvement; $MCIC_{imp}$) using the 80% specificity method), i.e., the cut point showing the best sensitivity for a response while still achieving at least 80% specificity, as previously recommended (in preference to the maximum accuracy method for improved comparability across different ROC analyses[29,30]. The same method was applied to identify the cut-off for the instrument scores at 6 months associated with "satisfaction with current state".

Analyses were carried out using SPSS (version 17, SPSS Inc, Chicago, IL, USA) and Medcalc (MedCalc Statistical Software, Mariakerke, Belgium). P values <0.05 were considered statistically significant. Unless otherwise stated, all data are presented as the mean and standard deviation (SD).

Results

Distribution of overall "treatment success" ratings

Table II shows the distribution of responses for the single items used to assess the "success" of surgery. The proportion of "good" outcomes, assessed with the GTO item, was slightly but

Table II
Distribution of responses in relation to the single item questions assessing the overall outcome of surgery for women, men and for the whole group together

Six-months' follow-up	Women		Men		*P* value	All	
	N	%	*N*	%		*N*	%
Change compared with preoperatively							
GTO							
Helped a lot (0)	17	30%	7	16%		24	24%
Helped (1)	20	36%	16	36%		36	36%
Helped only little (2)	15	27%	16	36%		31	31%
Did not help (3)	3	5%	5	11%		8	8%
Made things worse (4)	1	2%	0	0%		1	1%
Good outcome (0 and 1)	**37**	**66%**	**23**	**52%**	**0.162**	**60**	**60%**
Poor outcome (2–4)	**19**	**34%**	**21**	**48%**		**40**	**40%**
Total	56	100%	44	100%		100	100%
Acceptability of current state							
SSWB							
Very satisfied (0)	8	14%	5	12%		13	13%
Somewhat satisfied (1)	17	30%	11	25%		28	27%
Neither satisfied nor dissatisfied (2)	7	12%	8	18%		15	15%
Somewhat dissatisfied (3)	11	19%	12	27%		23	23%
Very dissatisfied (4)	14	25%	8	18%		22	22%
Acceptable state (0–2)	**32**	**56%**	**24**	**55%**	**0.873**	**56**	**55%**
Not acceptable state (3–4)	**25**	**44%**	**20**	**45%**		**45**	**45%**
Total	57	100%	44	100%		101	100%

Table III
Outcome scores before and 6 months after surgery in men and women

		N	Pre		6 months*	
			Mean	SD	Mean	SD
EQ-VAS (0 worst–100 best)						
Gender	Men	42	71.4	17.9	78.7	14.5
	Women	57	58.7	23.3	73.8	19.0
EQ-5D index (−0.59 worst–1.0 best)						
Gender	Men	44	0.700	0.220	0.805	0.214
	Women	57	0.476	0.229	0.707	0.260
OHS (0 worst–48 best)						
Gender	Men	42	36.3	7.1	39.8	6.9
	Women	56	29.2	7.7	38.0	7.5
WOMAC pain (0 best–100 worst)						
Gender	Men	43	31.8	24.3	15.2	16.6
	Women	57	46.7	21.0	17.4	17.8
WOMAC stiffness (0 best–100 worst)						
Gender	Men	43	25.6	27.0	15.5	19.0
	Women	57	42.2	28.2	19.5	22.2
WOMAC function (0 best–100 worst)						
Gender	Men	43	22.9	24.1	12.4	17.2
	Women	57	37.0	22.5	14.3	17.4
WOMAC-total score (0 best–100 worst)						
Gender	Men	43	25.0	23.6	13.2	16.8
	Women	57	39.4	21.2	15.5	17.2

* Significant effect of time ($P<0.001$) and gender × time interaction ($0.002 < P < 0.024$) for all the outcome measures.

non-significantly ($P = 0.16$) higher in the women (66%) than the men (52%).

Fifty-five percent of patients reported an acceptable symptom state, with no significant gender difference ($P = 0.87$).

There were no significant differences between surgical procedures in the proportions of patients with good and poor outcomes (arthroscopy, 60% good; mini-open, 61% good), or acceptable and not acceptable state (arthroscopy, 56% acceptable; mini-open, 55% acceptable).

Changes in outcome scores from preoperatively to 6 months' follow-up

For all outcome instruments, there was a significant ($P < 0.0001$) improvement in scores from before surgery to 6 months after surgery (Table III). The ceiling effects for WOMAC (12%) and OHS (4%) at 6 months were acceptable (<15%). The degree of improvement was not dependent on the type of surgery undertaken (arthroscopy vs mini-open; $P = 0.152–0.823$ for all dependent variables), and hence these data were pooled for further analysis. The gender × time interactions were significant ($0.002 < P < 0.024$) for all the dependent variables. Compared with the men, the women showed significantly ($P < 0.001$) worse baseline scores but similar 6-month scores, meaning that the improvements in their scores for all instruments were significantly ($P < 0.05$) greater than those for the men (Table III).

Figure 1 shows the effect sizes (SRMs) for each of the instruments for the "good" and "poor" outcome groups for the men and women separately. The effect sizes were greater for the women than for the men for all instruments. The separation between the effect sizes for the outcome groups (low SRM in the "poor" group; high SRM in the "good" group) was best (widest) for the two EQ-5D scales, the OHS and the WOMAC function subdomain. For the other WOMAC domains, the SRMs in the poor outcome group were not inconsiderable (especially for the women) and only slightly lower than those for the good outcome group.

Outcome scores at 6 months in relation to the acceptability of symptom status

For each instrument (EQ-5D, OHS, WOMAC and its subdomains) there was a highly significant ($P < 0.001$) difference between the mean scores for "acceptable" and "not acceptable" symptom states at 6 months (Table IV).

The cut-off (threshold) scores indicating an acceptable symptom state for each questionnaire are shown in Table V. The best estimates revealed that, in order to be considered an "acceptable" symptom state, scores at 6 months had to be approximately 80–95% of the best achievable score for the given scale.

MCIC scores for EQ-5D, OHS and WOMAC

Table VI shows the cut-off values for the change scores associated with a good global outcome. These represent the MCIC scores for improvement ($MCIC_{imp}$), and were equivalent to a change of 20–50% from baseline values, depending on the instrument in question.

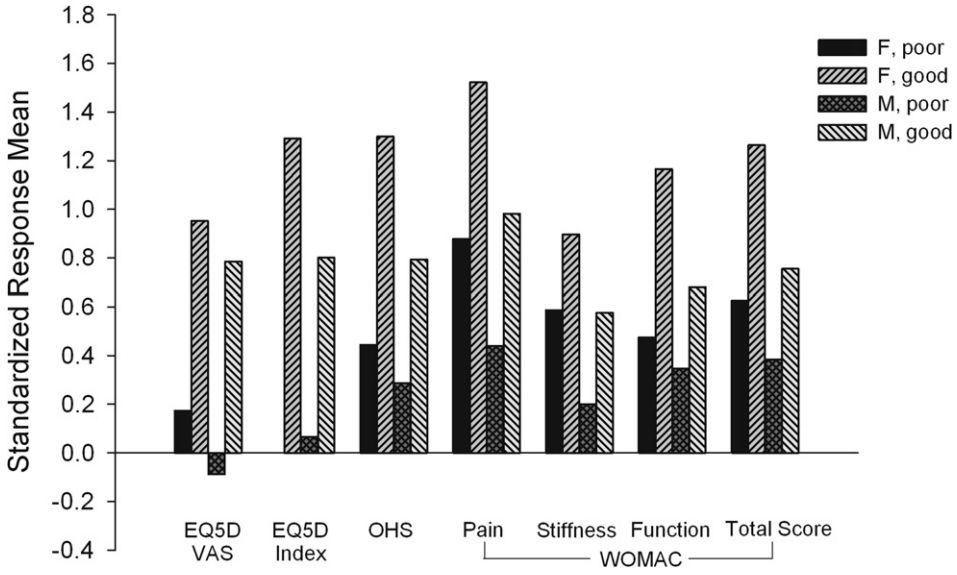

Fig. 1. Standardised response means for each outcome instrument, for the good and poor outcome groups (based on the dichotomised GTO), shown for men and women separately.

Table IV
Mean scores for each outcome instrument in relation to acceptability of symptom state as measured by SSWB

Instrument	Acceptable		Not acceptable		Mean difference* (95% CI)
	Mean	SD	Mean	SD	
EQ-VAS	84.4	10.4	65.7	18.4	18.7 (12.9–24.4)
EQ-5D index	0.881	0.143	0.587	0.248	0.294 (0.216–0.372)
OHS	42.5	4.6	34.4	7.4	8.1 (6.0–10.6)
WOMAC pain	9.6	11.4	26.1	20.3	16.5 (10.1–22.8)
WOMAC stiffness	9.5	13.3	29.3	24.6	19.9 (12.2–27.5)
WOMAC function	6.4	9.0	22.9	20.9	16.5 (10.3–22.6)
WOMAC-total score	7.4	9.2	24.2	20.4	16.7 (10.7–22.8)

* All mean differences $P < 0.001$.

Discussion

Summary of main findings

The present study examined the association between different expressions of "success" in relation to the early outcome of surgical treatment for FAI, quantifying both improvement ("feeling better"), measured on a transition scale and by the *change* in outcome scores, and acceptability of the current state ("feeling good"), measured on a Likert scale and by the *absolute* outcome scores at follow-up, as distinct clinically meaningful measures[16]. Although there was a significant improvement in scores measured with both the hip outcome tools and the generic instruments – to an extent similar to that previously reported at 2 years' follow-up in studies using one of the same instruments (WOMAC)[4,9,31,32] — only 60% patients reported notable clinical improvement and 55% deemed their current symptom state to be acceptable. It was no surprise that slightly fewer successes were recorded using the latter approach; previous studies have documented that feeling better does not always equate to feeling well[16]. This highlights the fact that improvements in outcome scores, even large and statistically significant, do not necessarily indicate an acceptable state. Our study also sought to provide guidance in the interpretation of individual questionnaire scores and we established that the threshold values for an acceptable symptom state were equivalent to 80–95% of the best achievable instrument score and that the change scores representing the minimal clinically important improvement for the hip outcome instruments were ≥6 for OHS and ≥21–28 for the various WOMAC domains (equivalent to improvements of 20–50% on the mean preoperative scores).

Table V
Results of ROC analysis to determine cut-off scores for each instrument indicating an acceptable state according to SSWB

Instrument	AUC*	95% CI	Cut-off value
EQ-VAS (0–100)	0.821	0.732–0.890	≥80
EQ-5D Index (−0.59–1.00)	0.840	0.754–0.905	≥0.682
OHS (0–48)	0.825	0.735–0.894	≥40
WOMAC pain (0–100)	0.793	0.701–0.867	≤8
WOMAC stiffness (0–100)	0.771	0.677–0.849	≤5
WOMAC function (0–100)	0.805	0.714–0.877	≤6
WOMAC-total score (0–100)	0.808	0.717–0.879	≤8

* All AUCs, $P < 0.001$.

Table VI
Results of ROC analysis to determine cut-off change scores for each instrument indicating a "good global outcome" according to GTO (i.e., the Minimum Clinically Important Change (MCIC))

Instrument	AUC*	95% CI	Absolute change scores cut-off value for a good outcome (MCIC)
EQ-VAS	0.752	0.654–0.833	≥15
EQ-5D index	0.780	0.686–0.857	≥0.16
OHS	0.762	0.665–0.843	≥6
WOMAC pain	0.735	0.636–0.818	≥28
WOMAC stiffness	0.626	0.523–0.721	≥25
WOMAC function	0.705	0.605–0.793	≥21
WOMAC-total score	0.721	0.622–0.807	≥22

* All AUCs, $P < 0.001$.

The external criteria of success

Patient-rated outcomes have been promoted heavily in recent years, but interpretation of the scores requires knowledge of what constitutes a meaningful (to the patient) score or change score after treatment. The provision of cut-off values improves the understandability and hence appeal of these outcome measures[33]. There is no gold standard methodology for estimating the score that constitutes the minimal clinically important improvement or the acceptable symptom state for an instrument. Determination of the $MCIC_{imp}$, involves either distribution-based methods relying only on the statistical characteristics (distribution) of the scores or anchor-based methods, which examine the relationship between change scores on the target instrument (e.g., WOMAC) and some independent measure of "worthwhile or important change"[34]. The latter follows more intuitively meaningful principles and are the recommended method of choice[35]. They require a valid, dichotomous external criterion to indicate "improvement" and this is commonly given by the patient's rating of the "global outcome of treatment"[36,37]. In the present study, the 5-category Likert scale was collapsed to produce a dichotomous outcome variable "good" and "poor". Although we do not suggest that this measure constitutes a definitive gold standard for assessing improvement, it is expected to reflect the most important changes to the individual patient elicited by the operation.

Acceptable symptom states after surgery for FAI

The calculation of thresholds associated with an "acceptable symptom state" yielded values of 80–95% of the best achievable score, depending on the scale in question. Few studies have previously investigated this concept, and none in patients with FAI, but these % values seem higher than those reported for some other musculoskeletal disorders. For patients with arthritis-related disorders, an expert panel in rheumatology proposed PASS values of 40/100 for the outcomes pain and function, i.e., patients were considered to be in an acceptable state if their score was below 40 on a 0–100 scale[38]. For patients with hip OA, a previous study reported PASS values of 35/100 mm for pain and 34/100 points for WOMAC function[14], which are even higher than the starting point for many of the FAI patients in the present study (especially the men). In contrast, the study of Falgarone *et al.* reported that three-quarters of all patients suffering from various musculoskeletal disorders considered their status as "acceptable" only when their pain level was less than 25/100[39] and Wells *et al.* suggested a comparable value of <20/100[35]. These values are closer to the levels reported in the present study but still indicate a slightly lesser "tolerance of symptoms" in our FAI patients. For more acute conditions – of which FAI may be more typical – the intensity of

symptoms considered "acceptable" is reportedly lower than in chronic disabling conditions, for which greater symptom levels and impairment appear to be tolerated[16,40]. Perhaps the relative youth and active lifestyle of FAI patients also renders them less tolerant of painful musculoskeletal conditions. Longer-term follow-up of the patients in the present study will allow us to examine whether the reported threshold values corresponding to an "acceptable state" are associated with a lesser need for further care or re-intervention; this should provide further validation of their clinical relevance.

MCIC scores for improvement after surgery for FAI

In the present study we demonstrated, as others have done (reviewed in van der Roer et al. and Lodhia et al.[9,15]), that mean post-treatment scores showed a significant improvement on pre-treatment values, some with relatively high effect sizes. The change-scores and effect sizes (SRM) were always higher in the "good" outcome group (see Fig. 1) but for some instruments/sub-domains (most notably WOMAC pain and stiffness) they were also moderately high in the "poor" outcome group, indicating that a number of patients who were not improved according to the global outcome criterion still showed improvement in these scores. Ideally, an outcome instrument should not indicate improvement when none has been deemed to occur, and this suggests a certain lack of specificity in the instrument. In order to better quantify the performance of the different outcome instruments in this respect, the ROC method was used. The ROC area under the curve (AUC) indicates the ability of the instrument change scores to distinguish between improved and not improved statuses, and in the present study AUCs ranged from 0.63 (WOMAC stiffness) to 0.78 (EQ-5D). The ROC method allowed us to establish that, for a given individual, a successful outcome would be indicated by a change in WOMAC-total score of at least 22 points, or by an increase in OHS equal to or greater than six points. As highlighted by others[41,42], it must be borne in mind that these $MCIC_{imp}$ values should not be considered as absolutely fixed values, but as approximations. The $MCIC_{imp}$ can vary depending on the patient group, their initial scores and the treatment under investigation, and also depending on the method and external criterion used for its determination[41,42]. The values reported in the current study are applicable to patients assessed 6 months after surgery for FAI and cannot necessarily be extended to patients with other conditions of the hip such as those undergoing total hip replacement.

The $MCIC_{imp}$ values reported here should be of assistance in the planning of clinical trials. For example, they can be used to determine the sample size required to detect a given difference in the percentage of patients achieving a $MCIC_{imp}$. Being able to express the results not only as mean values, but also as the proportion of patients meeting a given threshold for change, or achieving a certain score postoperatively, should better satisfy the demand for improved and simplified reporting of clinical trial results[33]; this, in turn, can be expected to facilitate interpretation and hence implementation of the findings of such trials in everyday clinical practice.

Limitations of the study

A number of limitations of the present study require mention. The follow-up was only 6 months, which might be considered by some to be rather short. However, since the study was intended to compare the degrees of success at a given time as judged by different outcome measures rather than report the final outcome of the surgical procedure *per se*, this period of follow-up was considered unproblematic. It allowed the early outcome of surgery to be assessed, once recovery and rehabilitation from the surgery itself were complete, at which point the main success (or otherwise) of the specific intervention should be evident. This is supported by the finding of similar scores for the WOMAC in the present study at 6 months (mean score of 86 when converted on a 0–100 scale with 100 indicating a good condition) and in previous studies with longer follow-ups of 2.0–2.5 years postoperatively (scores of 87 to 89[31,32]). Moreover, determination of the cut-off scores for relevant change/acceptable status are dependent on an accurate transition rating; intuitively, this is likely to be more reliable with a shorter recall period[43]. The uncertainty represented by the confidence intervals around the various cut-off values should be taken into consideration when interpreting the results of this study. Indeed, although the AUCs were significant, the lower bounds of their confidence intervals were, for some instruments and domains, below the 0.70 usually considered acceptable for accurately differentiating groups.

In presenting the $MCIC_{imp}$ values, we made no attempt to split up the patient population in relation to the type of surgical intervention, gender, severity of baseline status, etc., due to the relatively small size of the study group. These important subgroup analyses should be performed in future, once the systematic assessment of outcome in FAI patients has expanded sufficiently to become commonplace in clinical practice.

Conclusions

This study attempts to clarify what constitutes a good early outcome after surgery for FAI, from the perspective of the patient, in terms of the absolute values and improvements achieved on various outcome instruments. It shows that feeling better does not always equate to feeling well, and that improvements in outcome scores, even large, do not necessarily indicate acceptability of the current state. The $MCIC_{imp}$ values derived for the outcome instruments examined (EQ-5D index, EQ-VAS, OHS, WOMAC) should help in the interpretation of scores and score changes at the individual patient level in clinical trials and in daily clinical practice.

Author contributions

FMI, AFM, FDN, OH and ML were all responsible for the conception and design of the main study, of which this sub-study is part, and FMI, AFM, FDN and ML acquired funding for the project. OH and ML provided study patients. FMI coordinated all the practical work and acquisition of data. FMI and AFM performed the statistical analysis and interpreted the data. AFM created the first draft of the manuscript and FMI, FDN, OH and ML edited and revised it for important intellectual content. All authors read and approved the final manuscript.

FMI and AFM take responsibility for the integrity of the work as a whole, from inception to finished article.

Funding

The study was funded by the Schulthess Klinik Research Fund (internal funding) and by the SYNOS Foundation. None of the sponsors had any involvement in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Conflict of interests

All the authors declare that they have no financial and personal relationships with other people or organisations that could potentially and inappropriately influence (bias) their work and conclusions.

Acknowledgements

The authors would like to thank Sandra Spiss-Schatt, Mirjam Stoop and Murielle Jeker for their valuable assistance with the data collection. The study was funded by the Synos Foundation and the Schulthess Klinik Research Funds.

References

1. Leunig M, Ganz R. [FAI – concept and etiology]. Orthopade 2009;38:394–401.
2. Wenger DR, Kishan S, Pring ME. Impingement and childhood hip disease. J Pediatr Orthop B 2006;15:233–43.
3. Clohisy JC, St John LC, Schutz AL. Surgical treatment of femoroacetabular impingement: a systematic review of the literature. Clin Orthop Relat Res 2010;468:555–64.
4. Ng VY, Arora N, Best TM, Pan X, Ellis TJ. Efficacy of surgery for femoroacetabular impingement: a systematic review. Am J Sports Med 2010;38:2337–45.
5. Bedi A, Chen N, Robertson W, Kelly BT. The management of labral tears and femoroacetabular impingement of the hip in the young, active patient. Arthroscopy 2008;24:1135–45.
6. Lavigne M, Parvizi J, Beck M, Siebenrock KA, Ganz R, Leunig M. Anterior femoroacetabular impingement: part I. Techniques of joint preserving surgery. Clin Orthop Relat Res 2004:61–6.
7. Graves ML, Mast JW. Femoroacetabular impingement: do outcomes reliably improve with surgical dislocations? Clin Orthop Relat Res 2009;467:717–23.
8. Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: rationale for their use. J Bone Joint Surg Am 2009;91(Suppl 3):41–9.
9. Stevens MS, Legay DA, Glazebrook MA, Amirault D. The evidence for hip arthroscopy: grading the current indications. Arthroscopy 2010;26:1370–83.
10. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine 2006;31:578–82.
11. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris back pain questionnaire: part 1. Phys Ther 1998;78:1186–96.
12. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. Journal of Chronic Diseases 1986;39:897–906.
13. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, *et al*. Minimal clinically important differences: review of methods. J Rheumatol 2001;28:406–12.
14. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, *et al*. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Ann Rheum Dis 2005;64:34–7.
15. Lodhia P, Slobogean GP, Noonan VK, Gilbart MK. Patient-reported outcome instruments for femoroacetabular impingement and hip labral pathology: a systematic review of the clinimetric evidence. Arthroscopy 2010;27:279–86.
16. Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. Arthritis Rheum 2006;55:526–30.
17. Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53–72.
18. Greiner W, Graf v.d. Schulenburg J-M, Piercy J, Eds. (EuroQol) Plenary Meeting. Discussion Papers. Hannover Uni-Verlag: WItte 1999.
19. Prieto L, Sacristan JA. What is the value of social values? The uselessness of assessing health-related quality of life through preference measures. BMC Med Res Methodol 2004;4:10.
20. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 1996;78:185–90.
21. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and validity of the cross-culturally adapted German Oxford hip score. Clin Orthop Relat Res 2009;467:952–7.
22. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, *et al*. The use of the Oxford hip and knee scores. J Bone Joint Surg Br 2007;89:1010–4.
23. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
24. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, *et al*. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. Z Rheumatol 1996;55:40–9.
25. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, *et al*. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18:374–9.
26. Cherkin DC, Deyo RA, Street JH, Barlow W. Predicting poor outcomes for back pain seen in primary care using patients' own criteria. Spine (Phila Pa 1976) 1996;21:2900–7.
27. Stratford PW, Spadoni G, Kennedy D, Westaway MD, Alcock GK. Seven points to consider when investigating a measure's ability to detect change. Physiother Can 2002;54:16–24.
28. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol 2001;54:1204–17.
29. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. Arthritis Rheum 2009;61:313–20.
30. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: identifying reversible and irreversible components. Arthritis Rheum 2006;54:2784–92.
31. Ilizaliturri Jr VM, Nossa-Barrera JM, Acosta-Rodriguez E, Camacho-Galindo J. Arthroscopic treatment of femoroacetabular impingement secondary to paediatric hip disorders. J Bone Joint Surg Br 2007;89:1025–30.
32. Ilizaliturri Jr VM, Orozco-Rodriguez L, Acosta-Rodriguez E, Camacho-Galindo J. Arthroscopic treatment of cam-type femoroacetabular impingement: preliminary report at 2 years minimum follow-up. J Arthroplasty 2008;23:226–34.
33. Schunemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. Health Qual Life Outcomes 2006; 4:62.
34. Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. Health Qual Life Outcomes 2006;4:69.
35. Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, *et al*. Minimal clinically important difference module: summary, recommendations, and research agenda. J Rheumatol 2001;28:452–4.

36. Beurskens AJHM, de Vet HCW, Köke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996;65:71–6.
37. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395–407.
38. Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT, *et al*. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34: 1188–93.
39. Falgarone G, Zerkak D, Bauer C, Messow M, Dougados M. How to define a Minimal clinically individual state (MCIS) with pain VAS in daily practice for patients suffering from musculoskeletal disorders. Clin Exp Rheumatol 2005;23: 235–8.
40. Pham T, Tubach F. Patient acceptable symptomatic state (PASS). Joint Bone Spine 2009;76:321–3.
41. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, *et al*. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90–4.
42. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes 2006;4:70.
43. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002;55:900–8.